SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

(2)

1a. REPORT SECURITY CLASSIFICATION UNCI	1b. RESTRICTIVE MARKINGS
2a. SECUF NA	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; Distribution Unlimited
2b. DECL NA	
4. PERFOF UCD STATISTICS REPORT No. AFOSR-89-0386	5. MONITORING ORGANIZATION REPORT NUMBER(S) AFOSR-TR- 89-1882
6a. NAME OF PERFORMING ORGANIZATION University of California, Davis / 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION AFOSR/NM
6c. ADDRESS (City, State, and ZIP Code) Division of Statistics University of California Davis, CA 95616	7b. ADDRESS (City, State, and ZIP Code) Bldg. 410 Bolling AFB, DC 20332-6448
8a. NAME OF FUNDING/SPONSORING ORGANIZATION AFOSR / 8b. OFFICE SYMBOL (If applicable) NM	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER AFOSR-85-0268
8c. ADDRESS (City, State, and ZIP Code) BLDG. 410 Bolling AFB, DC 20332-1648	10. SOURCE OF FUNDING NUMBERS: PROGRAM ELEMENT NO. 61102F; PROJECT NO. 2304; TASK NO A5; WORK UNIT ACCESSION NO.

AD-A216 508

11. TITLE (Include Security Classification)
~~Final Report~~ Methods of Non-Parametric Inference

12. PERSONAL AUTHOR(S)
Jane-Ling Wang

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final Report	FROM 7/1/85 TO 6/30/89	July 25, 1989	14

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	
XXX	XXXXXXXXXXX	XX	

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Next page.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	UNCLASSIFIED

22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE (Include Area Code)	22c. OFFICE SYMBOL
Dr. C Baruch	(202) 767-4940	AFOSR/NM

DD Form 1473, JUN 86 Previous editions are obsolete. SECURITY CLASSIFICATION OF THIS PAGE

UNCLASSIFIED

90 01 04 144

ABSTRACT

This project deals with several nonparametric inference problems including two-sample tests, linear regression and estimation of distribution and related functions such as density and hazard rate functions. Estimators with desired aging properties were constructed for IFRA and NBU distribution functions respectively based on randomly censored data and shown to be $n^{1/2}$-equivalent to the product-limit estimator. Nonparametric maximum likelihood estimator and its strong consistency were also derived for an IFR distribution for unidentifiable cause-of-failure data. Local asymptotic properties (strong consistency, asymptotic normality and mean squared error) of the kernel density and hazard rate estimators were obtained via a recent i.i.d. representation of the product-limit estimator. The results on kernel estimates were applied to obtain point and interval estimates of the change-point of a hazard rate function. Several median type two-sample test procedures which allows early termination of the study were constructed. Some two-sample measures for differences of distribution functions were compared and used to analyze interdistribution income inequality. It is also demonstrated how to construct two-sample confidence intervals and testing procedures based on one-sample confidence intervals. An i.i.d. representation for the bivariate product-limit estimator was derived together with its bootstrap version to facilitate the linear regression problem for censored data.

TABLE OF CONTENTS

FINAL REPORT

Grant No. AFOSR-85-0268

Period of Support: July 1, 1985 to June 30, 1989.

Principal Investigator: Jane-Ling Wang

Project Title: 1. Nonparametric Estimation of Reliability and Related Functions (7/1/85 to 6/30/87)

2. Some Contributions to Nonparametric Inference and Reliability Theory (7/1/87 to 6/30/89)

A. Research Objectives:

Cumulative distribution function (or equivalently reliability function or survival function) and related functions, such as cumulative hazard function, density function and hazard rate function, play an important role in reliability theory. The objective of the first project is to obtain nonparametric estimators of such functions. The objective of the second project is to study several nonparametric inference problems with reliability implications. In both projects the observations on which the procedures are based, can be either i.i.d. or randomly censored. The research topics can be summarized in the following categories.

1. Nonparametric estimation of distribution (or equivalently, reliability) functions with aging properties.
2. Nonparametric estimation of density and hazard (or failure) rate functions
3. Nonparametric inference for unidentifiable cause-of-failure data

4. Two-sample inference procedures based on ranks
5. Linear regression with censored data

In the first topic we raised the question of how to estimate a distributing function based on i.i.d. observations or randomly right censored observations if the distribution is known to have certain aging property like it is increasing failure rate average (IFRA), new better than used (NBU) or new better than used in expectation (NBUE) and etc. The ultimate goal is to look for optimal estimator (in the sense of asymptotic minimaxity) for those distributions. General problems in the area of survival analysis is also of interest to us.

The second topic deals with the general problem of density and hazard rate estimations. While much research has been done for i.i.d. observations the study for random censored observations was still scattered. We intended to investigate properties of the kernel estimates of the density and hazard rate functions.

For competing risk data it may occur, both in medical and engineering contexts, that the cause of failure cannot be identified. We referred to such model as the unidentifiable random censorship (URC) model. The URC model differs from the conventional random censorship in the sense that, under the URC model one only gets to observe the smaller value of the true lifetime and censoring time without knowing whether the corresponding observation was censored or not. In order to avoid the unidentifiability problem and as an initiation we assumed that the censoring distribution is known in topic 3. General questions of estimating a life distribution under the URC model were discussed. If the life distribution is known to have a certain aging property the effort is then put on the search of maximum likelihood estimator (MLE) and its asymptotic behavior.

Topic four involves the classical two-sample problem of comparing two treatments or a new treatment with a control. An alternative representation for rank procedures was proposed which has the advantage of allowing possibly early termination of the study. Possibilities of adaptive or robust procedures will also be pursued.

The linear regression model encounters difficulties when the dependent variable or predictors are subject to random censoring. Several methods have been proposed in the literature to handle the case when only the dependent variable is subject to censoring. We proposed in topic 5 a new method of estimating the regression coefficients when both the dependent and independent variables are subject to censoring. Properties of the estimates will be explored and compared with existing procedures in a Monte Carlo study.

B. Status of Research

We shall now describe the progress and status of each topic.

1. My focus on estimation of distribution functions (topic 1) led to the development of an asymptotically minimax estimator for an IFRA distribution function. The new estimator is IFRA itself and is closer to the true distribution, in supnorm, than the sample distribution function (in i.i.d. case) or the Kaplan-Meier PL-estimator (in the censored case). For NBU distributions, the estimator of Boyles and Samaniego (1984) was shown, under mild regularity conditions to be $n^{1/2}$-equivalent to the sample distribution function of PL-estimator. Since the sample distribution or PL-estimator is asymptotically minimax for IFRA and NBU distributions, such optimality extends to our

estimators as well. The results for both the IFRA and NBU cases were included in Wang (1987b, publication #4). The two earlier technical reports on estimating star-shape and IFRA distribution respectively (i.i.d. case) were also published (Publication #3, 6) during the grant period. Other classes of aging distributions including NBUE, DMRL and HNBUE classes are also of interest and will be explored in the future.

2. For the estimation of density and hazard rate function (topic 2) under random censorship model. We first studied the local behavior of kernel type estimators. The strong consistency, law of iterated logarithm, asymptotic normality, expressions of mean squared error (MSE) and optimal rates of bandwidth for both the density and hazard rate estimates were established in Lo, Mack and Wang (1989, Publication #9). Our technique, which is facilitated by a recent result of Lo and Singh (1986) on an i.i.d. strong representation of the PL-estimator is relatively simple compared to the usual approaches based on strong embedding or counting process. It also provides a method to treat most local results in kernel density estimation and in particular, the MSE expressions and the optimal bandwidths.

The technique in this paper was also applied to another related problem which is the estimation of the change-point of the hazard rate of a certain item. The usual approach is to assume that such change takes place at one point when the constant hazard rate changes its value to another constant. Since actual changes may occur gradually and the usual change-point model can be approximated arbitrarily closely by smooth hazard rate, the concept of change-point is generalized to a smooth hazard rate. The usual change-point now corresponds to the location of an extremum of the derivative of the hazard rate, i.e., the point with the mose-rapid change. A nonparametric estimate for the point of the most-rapid change of a hazard rate is proposed in Müller and Wang (1988, submittd #3) as the location of an extremum of a kernel estimate of the derivative of the hazard rate. Consistency and limiting distribution of the estimator were obtained together with confidence intervals for both the derivatives of the hazard rate and the point of the most-rapid change. The confidence intervals were assessed in a Monte Carlo study and performed reasonably well for finite sample sizes.

3. In the occurence of unidentifiable cause-of-death (topic 3) we assume that (a) the life distribution corresponds to the cause-of-death of interest has increasing failure rate and (b) the censoring distribution, or in the competing-risk model the life distribution corresponds to death of all other causes, is known or can be estimated independently and quite accurately from earlier studies. Nonparametric maximum likelihood estimate and its consistency are derived in Mukerjee and Wang (1988, Submitted #1) using the framework of isotonic regressions. Several algorithms are given explicitely to compute the maximum likelihood estimator. Our next project in this area is to modify the assumptions (a) and (b) above to allow partially identifiable cause-of-death data and/or other shapes for the life distribution of interest.

4. Four papers were written during the grant period on the classical two-sample test of the effectiveness of a new treatment. In the first one (Gastwirth and Wang (1987, Publication #5)), an improved median type test is proposed and both small and large sample properties of the test statistic are presented. The methods are applied to data of an equal

employment case in which the minority fraction of the sample is invariably less than half so that the usual median test has zero power. The new test allows us to terminate the experiment earlier than most nonparametric test to reduce costs and is more powerful in general than the ordinary median test according to simulation results.

The second paper (Gastwirth and Wang (1988, Publication #7)) extends the control percentile test of Mathisen (1943) to accommodate censored data. Large sample distribution of the test statistic was derived and asymptotic efficiency of the tests were computed under the Koziol-Green model. As in the uncensored setting the efficacy of the censored version of the control median test equals that of the censored median tests. The relationship between the fraction of data that is censored and the efficacy of control percentile tests is explored numerically and the optimal percentile to use is shown to vary with the degree of censoring.

The third paper (Gastwirth, Nayak and Wang (1989. Publication #11)) has economics application and resulted while I was at the Wharton School of the University of Pennsylvania. In analyzing interdistributional income a variety of two-sample statistical measures have been used. Two recently introduced measures indicate a much larger secular change in the black-white income differential than the currently used measures. In order to understand this phenomenon both the theoretical properties of and empiric results obtained from five measures are given. It is shown that the new measures are more sensitive to the type of change that actually occurred than the usual measures which were designed to detect a shift in location. This paper is invited for a forthcoming special issue of the Journal of Econometrics.

The fourth paper (Wang and Hettmansperger (1988), submitted #2)) demonstrates how to construct two-sample confidence intervals and testing procedures based on one-sample intervals for randomly censored data.

Confidence intervals based on quantiles of the Kaplan-Meir product-limit estimator were derived for median survival times. Using these, two-sample tests and confidence intervals for the difference in median survival times are then developed based on the comparison of the one-sample confidence intervals. Several methods for choosing the confidence coefficients of the corresponding one-sample confidence intervals are developed under the shift model. The Pitman efficiencies of these two-sample tets are the same as that of the censored version of the median test. The procedures can also be applied to the Behrens-Fisher problem, proportional hazard model and accelerated failure-time model.

5. For the linear regression problem $Y = \alpha + \beta X + \varepsilon$ in topic 5, we assume the random design model that (X,Y) has a joint distribution F(x, y). New estimates $\hat{\beta}$ of β are proposed based on bivariate estimates of F(x, y). The properties of $\hat{\beta}$ thus depends on the properties of the bivariate estimates $\hat{F}(x, y)$. Several bivariate estimates are available for randomly censored bivariate data. Lo and Wang (1988, Publication #8) studied the bivariate Product-limit estimator of Campbell and Foldes (1982) and represent it as a mean of i.i.d. random variables. Large sample properties, e.g. asymptotic distribution and law of iterated logarithm, of the bivariate PL-process were derived as a result of the i.i.d. representation. The covariance structure of the limiting process was also given explicitly for the first time. Corresponding results were also derived for the bootstrap estimator which demonstrates the validity of the bootstrap procedure under the bivariate ramdom censorship model.

As for the slope estimator $\hat{\beta}$, some preliminary results on the consistency and asymptotic distribution are available. Monte Carlo study indicates that a resampling technique may perform well provided a good bivariate estimate is available. All of the existing bivariate estimates have some drawbacks one way or other and we are currently searching for a reasonable candidate. We are also running a large scale simulation to compare our procedures with several existing ones. Such a study is quite extensive and requires intensive computing effort. We hope to finish this project in the near future.

Other Research Supported by AFOSR

In addition to the above topics in my proposals, I also got involved in a joint project (with H.G. Müller) on dose-response curve. Bootstrap methods proposed by Efron in the late seventies have drawn central attention in the statistical community nowadays. An interesting question is whether the bootstrap confidence procedures out perform the standard maximum likelihood confidence procedures or not. Parametric bootstrap methods for the construction of confidence intervals for the effective dose at level α (ED α) under the probit model for the dose-response relationship are investigated.

The standard maximum likelihood confidence intervals and percentile, centered percentile, studentized, bias corrected and better bias corrected bootstrap methods are compared in a simulation with 1000 Monte Carlo runs and 1000 bootstrap samples. Among the bootstrap methods, studentized and centered percentile methods are found to behave unfavorable with respect to observed coverage probability, whereas the bias corrected and better bias corrected bootstrap sometimes improve on the maximum likelihood method. The maximum likelihood method yielded very mixed results, but in our simulation none of the currently available bootstrap methods improved uniformly on this standard method. All the above results are included in Müller and Wang (1988, Publication #10).

REFERENCES

1. Boyles, R.A. and Samaniego, F.J. (1984). Estimating a survival curve when new is better than used. Operation Research 32, 732-740.

2. Campbell, G. and Foldes, A. (1982). Large sample properties of nonparametric bivariate estimators with censored data. Proceedings of the International Colloquia on Nonparametric Inference, Budapest 1980, 23-28, North Holland.

3. Lo, S.H. and Singh, K. (1985). The product-limit estimator and the bootstrap: Some asymptotic representations. Probability Theory and Related Fields, 71, 455-465.

4. Mathisen, H.C. (1983). A method of testing the hypotheses that two samples are from the same population. Ann. Math. Statist. 14, 188-194.

* See also publication list.

PUBLICATIONS

1. Wang, J. (1985). Strong consistency of approximate maximum likelihood estimators with applications in nonparametrics. Ann. Statist., Vol. 13, 932-946.

2. Wang, J. (1986). Asymptotically minimax estimators for distributions with increasing failure rate. Ann. Statist., Vol. 14, 1113-1131.

3. Wang, J. (1987a). Estimation of distribution functions with increasing failure rate average. J. of Statist. Plann. and Inf., Vol. 16, 415-427.

4. Wang, J (1987b). Estimating IFRA and NBU survival curves based on censored data. Scand. J. of Statistics, Vol. 14, 199-210.

5. Gastwirth, J.L. and Wang, J. (1987). Nonparametric tests in small unbalanced samples: Application in employment - discrimination cases. Canadian J. of Statistics, Vol. 15, 339-348.

6. Wang, J. (1988) Optimal estimation of starshaped distribution functions. Statist. and Decision, Vol.6, 21-32.

7. Gastwirth, J.L. and Wang J. (1988). Control percentile test procedures for censored data. J. of Statist. Plann. and Inf., Vol. 18, 267-276.

8. Lo, S.H. and Wang, J. (1989). I.I.D. representations for the bivariate product limit estimators and the bootstrap versions. To appear in J. of Multivariate Analysis, Vol. 28, 211-226.

9. Lo, S.H., Mack, Y.P. and Wang, J. (1989). Density and hazard rate estimation for censored data via strong representation of the Kaplan-Meier estimator. Probability Theory and Related Fields, Vol. 80, 461-473.

10. Müller, H-G. and Wang, J. (1989). Bootstrap confidence intervals for effective doses in the probit model for dose-response data. To appear in Biometrical Journal.

11. Gastwirth, J.L. Nayak, T.K. and Wang, J. (1989). Statistical properties of measures of between group income differentials. Invited paper by the Journal of Econometrics.

MANUSCRIPTS SUBMITTED

1. Mukerjee, H. and Wang, J. (1988). Nonparametric maximum likelihood estimation of an increasing hazard rate for unidentifiable cause-of-death data.

2. Wang, J. and T.P. Hettmansperger (1988). Two-sample inference for median survival times based on one-sample procedures for censored survival data.

3. Müller, H-G and Wang, J. (1988). Nonparametric analysis of changes in hazard rates for censored survival data: An alternative to change-point models.

RESEARCH IN PROGRESS

1. Linear regression with censored data: Alternative approaches. Joint with S.H. Lo.

2. Hajek Projection for truncated and censored data. Joint with U. Uzunogullari.

3. Hazard rate estimation for truncated data. Joint with U. Uzuogullari.

4. Quantile representations for truncated data. Joint with U. Uzuogullari.

INTERACTIONS

1. "Two-sample inference procedures based on one-sample procedures for censored survival data" Institute of Medical Statistics, University of Erlangen-Nurnberg, West Germany, July 1988.

2. "Implications of the i.i.d. representation of the product-limit estimator" Department of Biostatistics, The Johns Hopkins University, December, 1987.

3. "New approaches to censored survival data" Department of Statistics, The Pennsylvania State University, September 1987.

4. "Bootstrap confidence intervals for ED α under the probit model" Institute of Medical Statistics, University of Erlangen-Nurnberg, West Germany, August 1987.

5. "Some new procedures for censored survival data" Department of Statistics, The Wharton School of the University of Pennsylvania, November 1986.

6. "Some new procedures for censored survival data" Washington Statistical Society, Washington, D.C., November 1986.

7. "Hazard rate estimation for censored data via strong representation of the Kaplan-Meier estimator" International Statistical Conference, Taipei, Taiwan, R.O.C., August 1986.

8. "I.I.D. representations for the bivariate product limit estimators and the bootstrap versions" Western regional meeting of the Institute of Mathematical Statistics, Seattle, Washington, July 1986.

9. "Density and hazard rate estimation for censored data" Department of Statistics, Rutgers University, April 1986.

CONTRIBUTED TALKS

1. "Two-sample inference procedures based on one-sample procedure for censored survival data", 14th International Biometric Conference, Namur, Belgium, July 1988.

2. "Optimal estimation of a NBU survival function based on censored data", 45th Session of the International Statistical Institute, Amsterdam, The Netherlands, August 1985.